

Measures of Language Proficiency from the Learner's Perspective*

Hossein Farhady

University of California, Los Angeles

University for Teacher Education, Tehran, Iran

Abstract

This paper examines the importance of learner characteristics in relation to learner performance on ESL tests. It is argued here that test taker characteristics are not included in the design of most ESL tests. Empirical evidence is provided to support the hypothesis that performance on various ESL tests is closely related to test takers' educational and language backgrounds. It is also argued that in order to account for those factors, and thus decrease test bias, the theoretical definition of language proficiency should be modified. Finally, some guidelines to dealing with test taker characteristics are suggested.

Introduction

For many years, language testers have focused on the theoretical and statistical dimensions of language testing. However, there is another dimension which has not received sufficient attention, namely, the people who take these tests.

As early as 1961, Carroll pointed out that the diversity of students' backgrounds and their previous preparations would make the task of language testers very demanding. Probably, because of the complexities involved in this issue, language testers have virtually ignored a very crucial factor in language testing, one that involves the characteristics of the test taker.

In this article, some of the theoretical and practical issues in **ESL** testing will be examined. More specifically, a) some of the inadequacies of various definitions of language proficiency with respect to test taker characteristics will be discussed, b) empirical evidence in support of the relationship between test taker characteristics and performance on language tests will be provided, and c) guidelines to dealing with test taker characteristics in language testing will be suggested.

Theoretical Problems

One of the principles of a scientific theory is the generation of hypotheses in which variables can be defined as clearly as possible. By a scientific theory, I mean a theory which can be substantiated and validated through empirical investigation. And by a hypothesis, I mean a tentative statement which predicts the relationship between two or more variables.

In language testing, various theories, including those of discrete-point (Lado, 1961), integrative (Carroll, 1961), pragmatic (Oller, 1978), and functional testing (Upshur,

1979; Farhady, 1980a) have been developed, and some have been supported by research results. These theories have generated numerous hypotheses involving such variables as the structure of language, instruments, test takers, and so forth. However, because of inadequate definitions, some of these hypotheses need to be reconsidered and the variables involved reexamined.

It should be noted that the lack of a scientific theory will weaken the external and/or internal validity of research and thus the validity of the results obtained from such research projects. Furthermore, if the hypotheses of the theory are poorly stated, they will result in poorly defined variables which will also make the interpretation of the results less defensible.

Language proficiency is one of the most poorly defined concepts in the field of language testing. Nevertheless, in spite of differing theoretical views as to its definition, a general issue on which many scholars seem to agree is that the focus of proficiency tests is on the students' ability to use language. Proficiency tests are supposed to be independent of the ways in which language is acquired. Brière (1972) points out that the parameters of language proficiency are not easy to identify. Acknowledging the complexities involved in the concept of language proficiency, Brière states:

The term 'proficiency' may be defined as: the degree of competence or the capability in a given language demonstrated by an individual at a given point in time independent of a specific textbook, chapter in the book, or pedagogical method (1972, p.332).

Such a complicated definition could very well result in vague hypotheses about language proficiency and language proficiency tests. They could be vague with respect to unspecified terms such as "competence", "capability", "demonstrated", and "individual". The term *competence* could refer to linguistic, socio-cultural, or other types of competence. The term *capability* could refer to the ability of the learner to recognize, comprehend, or produce language elements (or a combination of them). *Demonstration* of knowledge could be in either the written or the oral mode. Finally, the expression *individual* could refer to a language learner as listener, speaker, or both. These concepts should be clarified and their characteristics should be identified in order to develop explicit hypotheses.

Clark (1972) defines language proficiency as the language learner's ability

... to use language for real-life purposes without regard to the manner in which that competence was acquired. Thus, in proficiency testing, the frame of reference ... shifts from the classroom to the actual situation in which the language is used (p.5).

In this statement, another parameter is added to the function of language proficiency tests, namely, the use of language in real-life situations. That is, the statement includes all the complexities of previous definitions in addition to one more general concept, a 'real-life situation'.

Considering the difficulty of defining language proficiency, it is conceivable that the development and use of proficiency tests would involve more complex steps than

other types of language tests. This may be one of the factors that have slowed progress in testing language proficiency. Some scholars believe that language proficiency testing is the least advanced area in language testing (Clark, 1972). Although it is not an easy task to account for all aspects of language proficiency, it may be possible to, at least, clarify some of the ambiguous concepts involved in the definition of proficiency.

One of the major problems with the definitions above, and others as well, is that none of them includes test taker characteristics as a potential dimension in language testing. Theoreticians as well as practitioners have simply assumed that what the learners have learned and how they have learned it are irrelevant to language proficiency. This is, in my view, a gross and misleading assumption.

It has been demonstrated that learners from different educational backgrounds have certain performance profiles which indicate strengths and weaknesses in different language skills (Farhady, 1978; Hisama, 1977, 1978). Due to the educational policies in their home countries, students have differing views, conceptions, and perceptions of language tests as well as language instruction. Most of them differ in their relative needs for the use of language in their academic and social lives. The seriousness of the problem was observed twenty years ago by Carroll, who stated:

It is small wonder that a proposed external examination on English proficiency, designed for the testing of candidates from many countries and courses, will have to face the fact of profound differences in the kinds of preparations these candidates will have had (1961, p.314).

Of course, including learner variables in the definition of language proficiency will entail numerous problems, but they are worth considering. We claim that what the learner knows and how s/he has learned it can no longer be assumed to be irrelevant to the definition of language proficiency and that the parameters of language proficiency should be identified on the basis of learner as well as test characteristics.

There are many variables on which present tests are simply not designed to provide information. For example, factors such as learners' experience with test types, their weak and strong areas in various language skills, their knowledge of how and where to use language, the objectives of language courses they may be taking, and the relevance of these objectives to the students' academic as well as social lives, to name a few, have not been incorporated into the design of language proficiency testing.

Including each and any of these variables in a theory of language testing will require careful investigation and detailed examination of the nature of language tests. Testers should consider what the tests are measuring and what they should be measuring; what they expect a test to accomplish and what they should expect; which learner characteristics are included in language testing and which learner characteristics should be included. In short, the critical issue which deserves serious attention is what language testing is versus what it should be.

It is important to note that coining terminologies such as *general language proficiency*, of which no clear definition exists at this point, will not only make the problems go unnoticed but also misdirect research in the field. I do not intend to

review the arguments for and against such terminology because they have been frequently discussed in the literature (Oller, 1976, 1978; Spolsky, 1972, 1978; Clark, 1979; Hinofotis, 1976; Rand, 1972; Vollmer, 1979). What I intend to do is attempt to define some of the terms and propose research hypotheses which are empirically testable.

I have argued elsewhere (Farhady, 1980) that language proficiency is not a unidirectional phenomenon and that learners are not homogeneous in their proficiency in various language skills. Since the purpose of an instrument is to evaluate an attribute (which is multi-dimensional in this case) of people (who are heterogeneous in this case), language tests will have to serve multiple purposes in order to satisfy the requirements of an adequate information gathering process.

There is ample evidence in the literature which supports the multi-dimensionality of language behavior and the heterogeneity of test takers' abilities in different language skills. Research indicates that there are several factors underlying language proficiency tests (Oller & Hinofotis, 1980; Farhady, 1980a, b, c, and d; Vollmer, 1980, and many others). These factors could be identified as different skills such as listening comprehension, reading comprehension, speaking ability, and so forth. Furthermore, the heterogeneity of the language proficiency of test takers has also been demonstrated (Hisama, 1977; Farhady, 1978, 1979b). However, the hypothesis that learner variables are as important as other variables in language testing has not been fully investigated. Therefore, in the next section recent findings in support of this hypothesis are presented.

Empirical Evidence

The data presented here are part of a large-scale study designed to develop and validate functional language tests (Farhady, 1980). The experiment was carried out at UCLA with 800 incoming foreign students who took the UCLA English as a Second Language Placement Examination (ESLPE). The components of the Fall 1979 ESLPE are presented in Table 1.

TABLE 1
Number of Items in Each Subtest of the Fall '79 ESLPE

| Subtest | Number of Items |
|--------------------------|-----------------|
| Listening Comprehension: | |
| a. Visual | 10 |
| b. Written | 15 |
| Reading Comprehension | 15 |
| Grammar: | |
| a. Verb Forms | 20 |
| b. Prepositions | 15 |
| c. Others | 21 |
| Cloze | 25 |
| Dictation | 50 |
| Total | 171 |

One part of the data analysis examined the relationship between learner variables and learner performance. These variables included sex, university status (graduate vs. undergraduate), major field of study, and nationality. The analyses were conducted using standardized scores (T-scores) to eliminate the effect of the unequal number of items in the various subtests in the ESLPE.

1. Difference in Performance by Sex

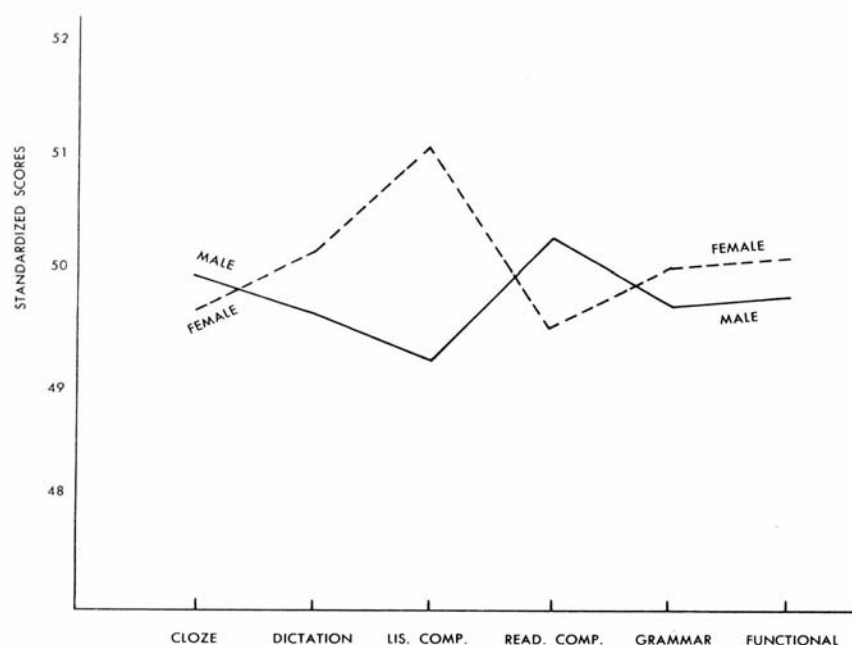
To compare the scores of male and female subjects on the study measures, t-tests were performed on their scores. The results, presented in Table 2, indicate no significant difference between male and female students in their performance on all but the listening comprehension subtest. Here, female students significantly outperformed male students. These results are illustrated in Figure 1.

TABLE 2
Standardized Descriptive Statistics for the Study
Measures by Sex

| Subtests | Male (N = 469) | | Female (N = 349) | | t |
|-------------|-------------------|-------|---------------------|-------|--------|
| | X | S | X | S | |
| Cloze | 49.78 | 9.85 | 49.70 | 10.15 | .13 |
| Dictation | 49.66 | 9.95 | 49.98 | 10.43 | -.44 |
| Lis. Comp. | 49.38 | 10.20 | 50.97 | 9.22 | -2.23* |
| Read. Comp. | 50.28 | 9.99 | 49.45 | 10.11 | 1.13 |
| Grammar | 49.77 | 9.74 | 49.88 | 10.42 | -.42 |
| Functional | 49.40 | 10.15 | 50.16 | 9.94 | -1.03 |

* $p < .05$

FIGURE 1
Difference Due to Sex in
Student Performance on Study Measures



2. Difference in Performance by University Status

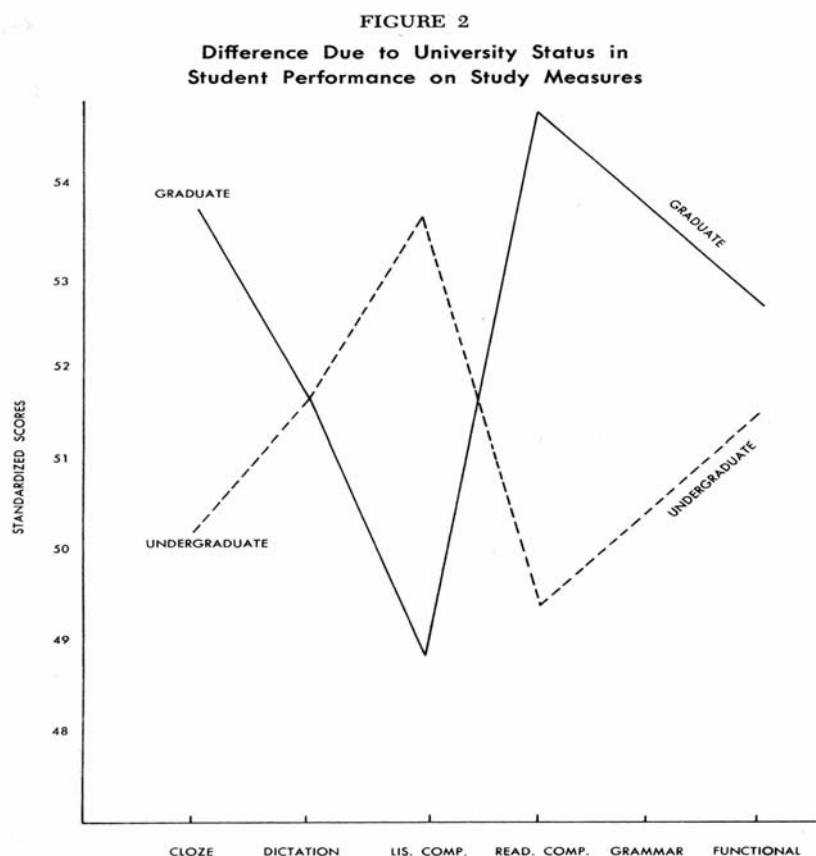
To determine the differential performance of the subjects according to their university status, t-tests were conducted on the scores of the graduate and undergraduate students. The results, reported in Table 3 and illustrated in Figure 2, indicate that these two groups of students performed differently on all but the dictation test. On the cloze, grammar, and reading subtests, graduate students significantly outperformed undergraduate students. On the listening comprehension test, however, the reverse happened.

TABLE 3
Standardized Descriptive Statistics for the Study
Measures by Status

| Subtests | Graduate (N = 284) | | Undergraduate (N = 378) | | t |
|-------------|-----------------------|------|----------------------------|------|---------|
| | X | S | X | S | |
| Cloze | 49.78 | 8.36 | 50.28 | 8.42 | 5.46** |
| Dictation | 52.25 | 8.04 | 51.78 | 7.25 | .63 |
| Lis. Comp. | 48.92 | 9.44 | 53.57 | 6.82 | -7.64** |
| Read. Comp. | 54.88 | 7.13 | 49.45 | 8.65 | 8.62** |
| Grammar | 53.85 | 7.24 | 50.57 | 8.67 | 5.16** |
| Functional | 53.21 | 8.26 | 50.63 | 8.68 | 3.86** |

** $p < .01$

Note: The scores for the Extension students (N = 158) are not presented here.



The higher performance of graduate students on reading and grammar subtests may be due to their extensive practice in these areas, whereas undergraduate students have not had enough opportunities to master grammatical rules or read much material in English. On the other hand, the higher performance of undergraduate students on listening comprehension could be due to various factors such as their age (undergraduate students tend to be younger than graduates) or length of residence in English speaking communities. It could also be due to recent changes in educational systems in foreign countries which emphasize oral-aural skills more than traditional systems did.

3. Difference in Performance by Major Field of Study

Separate one-way ANOVAs were conducted to investigate the difference in examinees' scores on the study measures due to their major fields of study. All university major fields were categorized into eight groups presented in Table 4.

The eighth group included 152 subjects whose major fields were not determined and thus were not included in the analyses. Descriptive statistics for the subtest scores by major field of study are presented in Table 5 and illustrated in Figure 3.

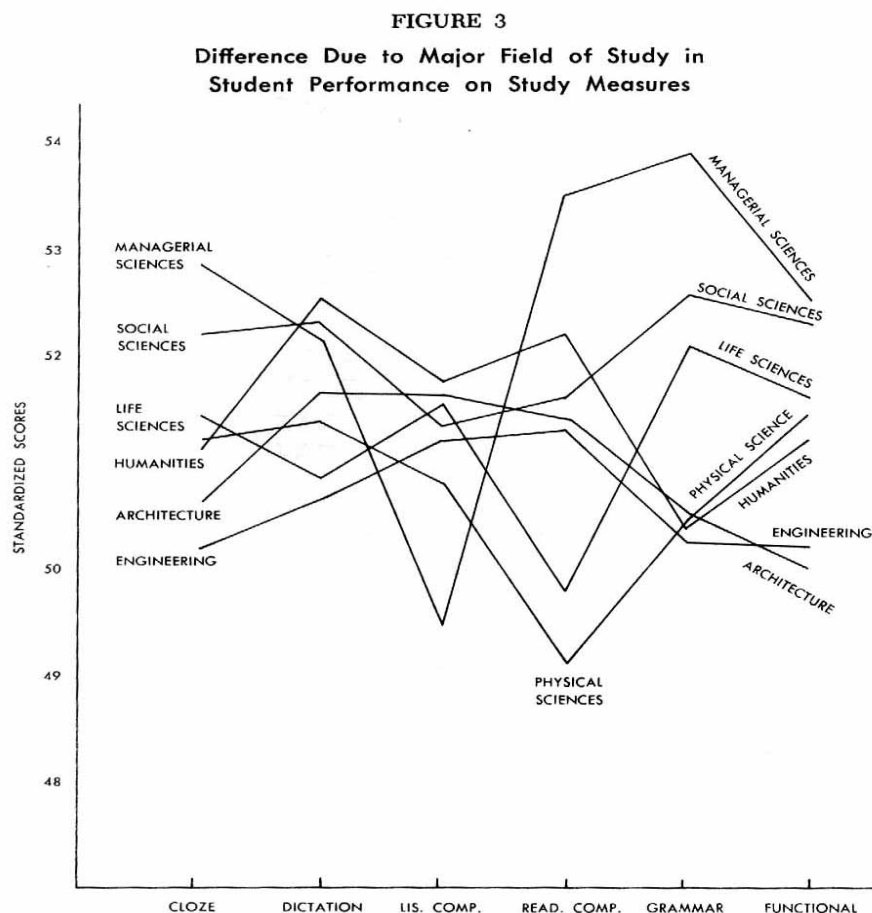
TABLE 4
Classification of Major Fields of Study

| Major Fields of Study | |
|-----------------------|---|
| 1. | Physical Sciences: (Physics, Chemistry . . .) |
| 2. | Engineering and Applied Sciences: (Math, Computer Science . . .) |
| 3. | Life Sciences: (Biology, Public Health, Medicine . . .) |
| 4. | Social Sciences: (Economics, Education, Psychology, Languages, Anthropology . . .) |
| 5. | Management: (Business, accounting . . .) |
| 6. | Architecture: (Urban planning, Geography, Design . . .) |
| 7. | Humanities: (Music, Art, Theater . . .) |
| 8. | Undeclared: (Extension) |

TABLE 5
Standardized Descriptive Statistics for
the Study Measures by Major Fields of Study at UCLA

| Major | N | Cloze | | Dictation | | Lis. Comp. | | Read. Comp. | | Grammar | | Functional | |
|---------------------------|-----|-----------|-------|-----------|-------|------------|-------|-------------|-------|-----------|-------|------------|-------|
| | | \bar{X} | S | \bar{X} | S | \bar{X} | S | \bar{X} | S | \bar{X} | S | \bar{X} | S |
| 1. Physical S. | 75 | 51.14 | 8.67 | 51.48 | 6.75 | 50.87 | 10.12 | 49.04 | 9.88 | 50.44 | 7.66 | 51.48 | 6.28 |
| 2. Engineering S. | 222 | 50.24 | 7.85 | 50.82 | 8.43 | 51.12 | 7.80 | 51.24 | 7.86 | 50.29 | 8.52 | 50.25 | 7.61 |
| 3. Life S. | 78 | 51.48 | 9.52 | 50.89 | 9.07 | 51.61 | 10.04 | 49.97 | 9.32 | 52.06 | 8.81 | 51.59 | 7.83 |
| 4. Social S. | 146 | 52.24 | 8.80 | 52.27 | 8.30 | 51.45 | 8.71 | 51.74 | 9.06 | 52.67 | 8.12 | 52.35 | 7.75 |
| 5. Management | 73 | 52.95 | 10.23 | 52.17 | 7.56 | 49.60 | 9.32 | 53.64 | 10.02 | 53.95 | 9.14 | 52.66 | 10.95 |
| 6. Architecture | 39 | 50.68 | 9.41 | 51.87 | 7.23 | 51.67 | 8.00 | 51.42 | 7.65 | 50.47 | 7.16 | 50.09 | 10.05 |
| 7. Humanities | 34 | 51.17 | 9.15 | 52.63 | 9.39 | 51.82 | 8.84 | 52.20 | 9.03 | 50.49 | 10.97 | 51.28 | 7.79 |
| 8. Undeclared (Extension) | 152 | 44.06 | 12.35 | 43.15 | 13.74 | 44.81 | 13.07 | 44.24 | 13.07 | 43.33 | 12.78 | 44.73 | 12.34 |

Measures of Proficiency



The results of the ANOVAs, reported in Table 6, indicate that students from different major fields of study performed significantly different on reading, grammar, and functional subtests. These differences should be suggestive rather than definitive because they could be due to procedures used to classify various major fields. A detailed classification of the major fields and multivariate analyses would be necessary to validate these results.

TABLE 6
The Results of ANOVA on the Study Measures
According to Major Field of Study

| Subtests | SSB | MSB (df = 5) | SSW | MSW (df = 628) | F |
|-------------|------|-----------------|-------|-------------------|--------|
| Cloze | 599 | 119 | 48403 | 77 | 1.55 |
| Dictation | 255 | 51 | 41588 | 66 | .77 |
| Lis. Comp. | 220 | 44 | 48832 | 77 | .57 |
| Read. Comp. | 954 | 190 | 49084 | 78 | 2.44* |
| Grammar | 1113 | 222 | 44019 | 70 | 3.18** |
| Functional | 1031 | 206 | 49672 | 79 | 2.61* |

* $p < .05$

** $p < .01$

4. Difference in Performance by Nationality

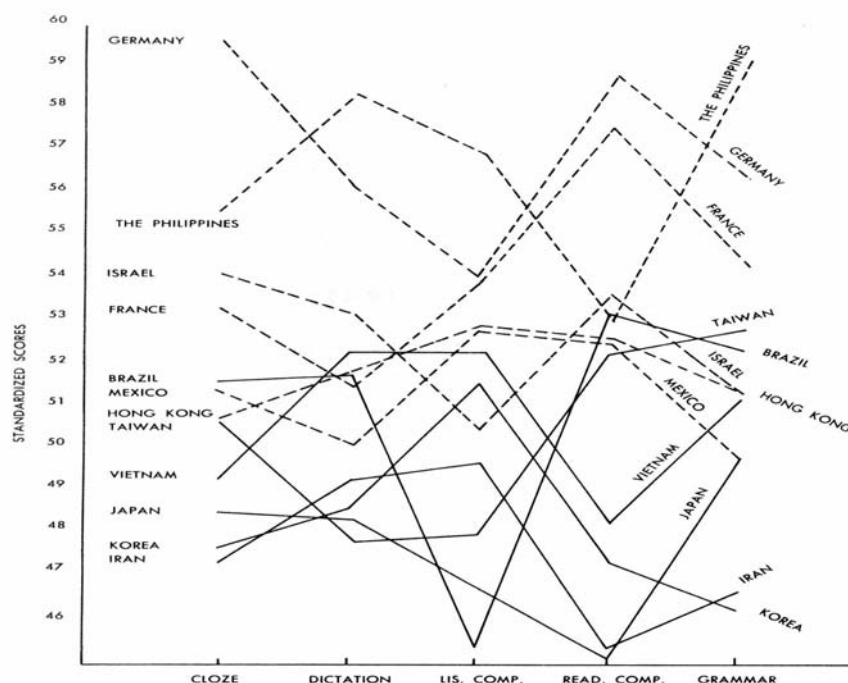
The findings of previous research on the ESLPE (Farhady, 1978, 1979b) suggested that there was a significant relationship between the examinees' nationalities and their performance profile. It was hypothesized that the relationship was due to different educational policies in different countries. That is, in some countries, ESL instruction may emphasize one language skill more than others. Therefore, the data in the present study were analyzed to either support or reject the validity of previous findings. Twelve countries (those with more than 15 students taking the ESLPE) were included in the analysis. The descriptive statistics are presented in Table 7 and illustrated in Figure 4.

TABLE 7
Standardized Descriptive Statistics for
the Study Measures by Nationality

| Nationality | N | Cloze | | Dictation | | Lis. Comp. | | Read. Comp. | | Grammar | | Functional | |
|---------------------|-----|-----------|-------|-----------|-------|------------|-------|-------------|-------|-----------|-------|------------|-------|
| | | \bar{X} | s | \bar{X} | s | \bar{X} | s | \bar{X} | s | \bar{X} | s | \bar{X} | s |
| 1. Iran | 140 | 47.34 | 10.73 | 49.17 | 10.43 | 49.58 | 10.80 | 45.33 | 10.80 | 50.44 | 7.66 | 51.48 | 6.28 |
| 2. Taiwan | 127 | 50.66 | 7.02 | 47.98 | 7.99 | 47.99 | 7.17 | 52.02 | 7.17 | 50.29 | 8.52 | 50.25 | 7.61 |
| 3. Korea | 117 | 47.62 | 9.11 | 48.64 | 8.86 | 51.57 | 8.58 | 47.43 | 8.61 | 52.06 | 8.81 | 51.59 | 7.83 |
| 4. Japan | 69 | 48.71 | 9.12 | 48.34 | 8.31 | 45.03 | 8.52 | 49.62 | 10.12 | 52.67 | 8.12 | 52.35 | 7.75 |
| 5. Vietnam | 32 | 49.28 | 9.23 | 52.17 | 6.62 | 52.39 | 5.46 | 48.26 | 10.33 | 53.95 | 9.14 | 52.66 | 10.95 |
| 6. Mexico | 26 | 51.54 | 13.00 | 50.02 | 15.99 | 52.89 | 12.10 | 52.59 | 10.66 | 50.47 | 7.16 | 50.09 | 10.05 |
| 7. France | 24 | 54.10 | 8.24 | 53.31 | 7.45 | 50.68 | 8.88 | 53.69 | 8.97 | 50.49 | 10.97 | 51.28 | 7.79 |
| 8. Hong Kong | 23 | 50.77 | 8.31 | 51.71 | 7.06 | 52.94 | 5.86 | 52.73 | 7.14 | 43.33 | 12.78 | 44.73 | 12.34 |
| 9. Israel | 22 | 53.38 | 9.23 | 51.54 | 12.46 | 54.17 | 5.25 | 57.56 | 4.23 | 54.32 | 7.40 | 56.60 | 6.18 |
| 10. Germany | 19 | 59.73 | 4.82 | 56.21 | 3.59 | 54.03 | 7.21 | 58.83 | 4.10 | 56.34 | 5.58 | 58.16 | 5.94 |
| 11. Brazil | 17 | 51.78 | 7.24 | 51.88 | 8.23 | 45.43 | 14.29 | 53.69 | 8.99 | 52.46 | 9.97 | 50.73 | 9.90 |
| 12. The Philippines | 16 | 55.72 | 5.81 | 58.30 | 3.88 | 57.05 | 3.95 | 53.09 | 7.12 | 58.97 | 3.82 | 59.51 | 6.02 |

Measures of Proficiency

FIGURE 4
Difference Due to Nationality in
Student Performance on Study Measures



The results of the ANOVAs, reported in Table 8, support the findings of the previous investigations. The data indicate that nationality is a strong factor in relation to the students' degree of language proficiency in various tests. For all subtests, the F values are significant at the .01 level indicating that students from different countries performed differently on the various language skill tests in this study.

TABLE 8
The Results of ANOVA on the Study Measures
According to Nationality

| Subtests | SSB | MSB (df = 11) | SSW | MSW (df = 620) | F |
|-------------|------|------------------|-------|-------------------|--------|
| Cloze | 4917 | 447 | 51026 | 82 | 5.43** |
| Dictation | 3420 | 311 | 51715 | 83 | 3.73* |
| Lis. Comp. | 4976 | 452 | 55344 | 89 | 5.07** |
| Read. Comp. | 8231 | 748 | 50170 | 81 | 9.24** |
| Grammar | 6884 | 626 | 52044 | 84 | 7.45** |

* $p < .05$

** $p < .01$

Note: To save space, the decimal values are not reported in the ANOVA tables.

The results reported on here do not support the findings of previous research on the ESLPE. Sanneh (1977) reported that factors such as students' sex, status, major field of study, and nationality did not have a significant relationship to student performance on the ESLPE. However, the data in this study suggest that these factors are significantly related to student performance on the ESLPE. The discrepancies between the results of this study and those reported on by Sanneh could be due to the modifications in the content of the ESLPE or changes in the student population and their proficiency patterns, or both.

Discussion

The differential performance of the subjects from different countries suggests that students coming to the university do not have similar training with respect to different language skills. It is not clear at this point why in some countries the focus of instruction is on one skill rather than another. What is clear, however, is that these differences do exist and should be dealt with somehow. Identifying and/or controlling the instructional factors, which are related to variations in incoming students' performance, is probably not within the power of language testers because they do not prescribe English language programs around the world. The important point, however, is that such differences, which may influence the efficiency of ESL testing and teaching at the universities, should be accounted for by modifying the test content, instructional objectives, or both.

The preponderance of evidence presented in this paper suggests that learner characteristics have a strong relationship with learner performance. Therefore, ignoring all these factors, simply defining language proficiency as a concept independent of learner variables seems unjustified. Since there was no significant

difference in the total scores of language groups on the ESLPE, it could be assumed that test taker characteristics were the factors which resulted in different performance patterns. Thus, if some of these variables could be incorporated in the design of testing programs, it would be a step in the right direction.

No matter what the purpose of the test may be, learner variables will definitely influence test scores in one way or another. That is, placement, selection, aptitude, proficiency, and other uses of language test scores will be sensitive to those who take the test. Furthermore, performance on discrete point, integrative, functional, or other types of tests will also be sensitive to those who take the test. Therefore, considering these factors in language test designs seems warranted.

Suggestions

Most existing ESL tests, which do not include learner variables in the data analysis and in the interpretation of test scores, have probably failed to assess learners' language abilities accurately. Consequently, a number of language learners may have been misplaced in ESL classes or denied admission to universities because of their low scores on language tests. These potential misjudgments could frustrate students or reduce their motivation. To avoid some of these undesirable consequences, the steps suggested below may be useful.

There are a number of factors that could contribute to improving ESL testing processes. These factors could be classified into three major categories: psychometric, typologic, and learner factors.

The psychometric factors involve the reliability and validity of the tests. Almost all language tests have been reported to have reasonably high internal consistency (alpha) coefficients. However, except for concurrent validity reports, few language tests have been examined for their content and construct validity. Almost all language tests seem to consist of randomly selected items of non-specified content materials. This is one area that could be improved. Administrators in ESL programs should be willing to examine the correspondence between test items and instructional objectives in order to increase the content validity of their tests.

For example, for a program intended to prepare competent speakers, a multiple choice test of grammar, no matter how reliable and valid it may be, will be neither sufficient nor appropriate. In other words, a given test will not be suitable for all examinees in all programs. There should be a direct relationship between the student, the test, and the instructional objectives (Carroll, 1980).

The typologic factors refer to the types of tests being used. Discrete-point, integrative, and functional tests all have their own advantages and disadvantages. It is not safe to assume that either one, or a combination of all for that matter, will constitute a perfect measure of language proficiency.

If the characteristics of these tests were known (i.e., item specifications were clearly developed) and if ESL programs were designed with specified objectives (i.e., instructional objectives were clearly developed), then, there would be no room to compromise because it would be easy to decide on an appropriate test format

regardless of learner variables. However, item specifications for most of the tests have not been developed and the tests seem to serve very similar purposes and provide similar information about examinees' performance. Nor have the objectives of instructional programs been clearly defined. Most ESL programs follow a similar general English instruction format. Therefore, a definite decision cannot be made about the format of the tests at present.

The learner factors involve variables related to the people who take the test. This is the area which needs careful reexamination.

It has been demonstrated that given a reliable and seemingly valid test, different types of tests provide similar information on the examinee's proficiency (Farhady, 1979a, 1979b). However, students with different backgrounds tend to perform differently on various language tasks. There seem to be two solutions to this problem, a short-term solution and a long-term solution.

A short-term solution, which is reasonably easy and immediately applicable, calls for a detailed analysis of test scores with a multidimensional design, i.e., including the learner variables mentioned above. The variables which yield significant differences among examinees will be selected for further exploration. Through various multiple regression analyses, adjustment formulas could be developed. Finally, examinees' total scores would be computed on the basis of regression coefficients associated with every subskill score. Such total scores will be unbiased, at least statistically, with respect to learner variables and test formats.

A long-term solution may apply for all language programs. That is, any program would start with a detailed analysis of learner needs. Then, the instructional objectives of the program would be established and achievement criteria would be determined. Finally, the testing procedures would be similar to those suggested by Carroll (1980). According to Carroll, there would be a two-phase testing program for language learners. The first phase would assess the learners' general and what Hinofotis (1981) calls *base level* English. Those who receive a satisfactory score on this test would go on to the next phase and take a test which is developed on the basis of careful analyses of learner needs. This test would assess a selective functional proficiency of the learners in various academic areas (Farhady, 1981). This means that learners from different educational disciplines might be required to take different tests.

In this manner, neither learner variables nor format factors can interfere with the decisions made on the basis of test scores because the test is devised to measure the elements which are necessary for the group. This is where tests of English for specific programs could be developed and utilized effectively.

Conclusions

Making conclusive statements about accurate assessment of learner's language proficiency is premature at this point. However, I believe that language testing is at a critical stage of evolution. The trend is shifting, on the one hand, from testing linguistic elements to testing communicative functions, and on the other hand, from using one all-purpose language test to specific and discipline-oriented measures. Therefore, it seems crucial to consider as many variables as possible and take them

into account in designing language tests. Without careful planning, the diversification of language tests will not be as effective as it should be.

* This is the revised version of the paper printed in *TESOL Quarterly* (1982), 16(1).

Bibliography

- Briere, E.J. (1972). Are we really measuring proficiency with our foreign language tests? In H.B. Allen & R.N. Campbell (eds.), *Teaching English as a second language: A book of readings* (2nd ed.). 1972. New York: McGraw-Hill Book Company.
- Carroll, J.B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing English proficiency of foreign students*. Washington, D.C.: Center for Applied Linguistics. Also reprinted in H.B. Allen & R.N. Campbell (eds.), *Teaching English as a second language: A book of readings* (2nd ed.). 1972. New York: McGraw-Hill Hook Company.
- ____ (1980). *Testing communicative performance*. Oxford: OUP.
- ____ (ed.) (1978). *Direct testing of speaking proficiency: Theory and application*. Princeton, NJ: Educational Testing Service.
- ____ (1979). Direct vs. semi-direct tests of speaking ability. In E.J. Briere & F. Hinofotis (eds.), *New concepts in language testing: Some recent studies*. Washington, D.C.: TESOL.
- Farhady, H. (1978). *The differential performance of foreign students on discrete-point and integrative tests*, MA thesis, University of California, Los Angeles.
- ____ (1979a). The disjunctive fallacy between discrete-point and integrative tests. *TESOL Quarterly*, 13(3).
- ____ (1979b). Test bias in language placement examinations. In C. Yorio & J. Schachter (eds.), *On TESOL '79*. Washington, D.C.: TESOL.
- ____ (1980). *Justification, development, and validation of functional language testing*, PhD dissertation, University of California, Los Angeles.
- ____ (1980b). *Rationalization and development of functional testing*. Paper presented in the 14th annual TESOL convention, San Francisco, CA.
- ____ (1980c). On the plausibility of the unitary language proficiency factor. In W.J. Oller, Jr. (ed.), *Issues in language testing research*. Rowley, Mass.: Newbury House Publishers, Inc.
- ____ (1980d). New directions for ESL proficiency testing: Language proficiency factor. In W.J. Oller, Jr. (ed.), *Issues in language testing research*. Rowley, Mass.: Newbury House Publishers, Inc.
- ____ (1981). *Testing functional ESL proficiency in ESP contexts*. Unpublished manuscript, UCLA.
- Hinofotis, F.B. (1976). *An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language*, PhD dissertation,

Southern Illinois University.

_____. (1981). *Perspectives on language testing: Past, present, future*. Nagoya Gakuin University Roundtable on languages, linguistics, and literature. Nagoya Gakuin University, Seta, Aichi, Japan.

Hisama, K. (1977). Patterns in various ESOL proficiency test scores by native language and proficiency levels. *Occasional Papers in Linguistics*. Southern Illinois University.

_____. (1978). An analysis of various ESL proficiency tests. In W.J. Oller, Jr. & K. Perkins (eds.), *Research in language testing*. Rowley, Mass.: Newbury House Publishers, Inc.

Lado, R. (1960). English language testing: Problems of validity and administration. *English Language Teaching*, 14.

_____. (1961). *Language testing: The construction and use of foreign language tests*. New York: McGraw-Hill Book Company.

Oller, J.W., Jr. (1976). Evidence for a general language proficiency factor: An expectancy grammar. *Die Neueren Sprachen*, 2, 165-171.

_____. (1978). Pragmatics and language testing. In B. Spolsky (ed.), *Advances in language testing, Series 2*. Arlington, VA: Center for Applied Linguistics.

Oller J.W., Jr. & F. Hinofotis (1980). Two mutually exclusive hypotheses about second language ability: Indivisible or partially divisible competence. In J.W. Oller & K. Perkins (eds.), *Research in language testing*. Rowley, Mass.: Newbury House Publishers, Inc.

Rand, E. (1972). Integrative and discrete tests at UCLA. In J. Povey (ed.), *Workpapers in TESL*. University of California, Los Angeles.

Sanneh, O. (1977). *A computer analysis of the English as a second language placement examination (ESLPE) results, 1974, '75, '76*. Unpublished paper. Office of International Students and Scholars, University of California, Los Angeles.

Spolsky, B. (ed.) (1972). *Advances in language testing, Series 2*. Arlington, VA: Center for Applied Linguistics.

_____. (1978). *Approaches to language testing*. Arlington, VA: Center for Applied linguistics.

Upshur, J.A. (1979). Functional proficiency theory and a research role for language tests. In E.J. Briere & F. Hinofotis (eds.), *New concepts in language testing: Some recent studies*. Washington, D.C.: TESOL.

Vollmer, H.J. (1979). *Why are we interested in general language proficiency?* Paper

presented at the 79th German International Symposium in Language testing in Hurth.

____ (1980). *Competing hypotheses about second language ability: A plea for caution*. Berlin: Osnabruck.